

# Achieving the Scaling Law of SNR-Monitoring in Dynamic Wireless Networks

Hongyi Yao<sup>1</sup>, Xiaohang Li<sup>2</sup>, Soung Chang Liew<sup>3</sup>

<sup>1</sup> California Institute of Technology <sup>2</sup> Purdue University <sup>3</sup> The Chinese University of Hong Kong

## Abstract

The characteristics of wireless communication channels may vary with time due to fading, environmental changes and movement of mobile wireless devices. Tracking and estimating channel gains of wireless channels is therefore a fundamentally important element of many wireless communication systems. In particular, the receivers in many wireless networks need to estimate the channel gains by means of a training sequence. This paper studies the *scaling law* (on the network size) of the overhead for channel gain monitoring in wireless networks. We first investigate the scenario in which a receiver needs to track the channel gains with respect to multiple transmitters. To be concrete, suppose that there are  $n$  transmitters, and that in the current round of channel-gain estimation,  $k \leq n$  channels suffer significant variations since the last round. We prove that “ $\Theta(k \log((n+1)/k))$  time slots” is the *minimum* overhead needed to catch up with the  $k$  varied channels. Here a time slot equals one symbol duration. At the same time, we propose a *novel* channel-gain monitoring scheme named ADMOT to achieve the overhead *lower-bound*. ADMOT leverages recent advances in compressive sensing in signal processing and interference processing in wireless communication, to enable the receiver to estimate all  $n$  channels in a *reliable* and *computationally efficient* manner within  $\mathcal{O}(k \log((n+1)/k))$  time slots. To our best knowledge, all previous channel-tracking schemes require  $\Theta(n)$  time slots regardless of  $k$ . Note that based on above results for single receiver scenario, the scaling law of general setting is achieved in which there are *multiple* transmitters, relay nodes and receivers.

Index terms: Wireless Network, Scaling Law, Channel Gain Estimation, Compressive Sensing.

## I. INTRODUCTION

The knowledge of channel gains is often needed in the design of high performance communication schemes [1], [2], [3], [4], [5]. In practice, due to fading, transmit power instability, environmental changes and movement of mobile wireless devices, the channel gains vary with time. Tracking and estimating channel gains of wireless channels is therefore fundamentally important [6], [7], [8], [9], [10], [11], [12].

An issue of interest is how to reduce the overhead of channel-gain estimation. On the one hand, if between two rounds of channel-gain estimation, the channels have varied significantly, then communication reliability will be

jeopardized [4], [5], [9]. On the other hand, if the frequency of channel-gain estimation is high, the overhead will also be high [1], [6], [13]. Our approach is predicated on reducing the overhead in each round, while maintaining high accuracy.

We first consider the case in which a receiver needs to estimate the channel gains from  $n$  transmitters [1], [2]. As a mental picture, the reader could imagine the receiver to be a base station, and the transmitters to be mobile devices. To achieve reliable bit-error-rate (BER), the frequency of estimation should be high enough [1]. Then it is likely that only a few of the  $n$  channels have suffered appreciable changes since the last estimation. We make use of the techniques of compressive sensing and interference signal processing to reduce the time needed to perform the estimation in each round. We answer the following question:

Suppose that in the current round, there are at most  $k \leq n$  channels suffering from appreciable channel gain variations. Given a target reliability for channel-gain estimation, what is the minimum overhead needed?

We answer this question by analysis and construction. We prove that the minimum number time slots needed for estimation is  $\Theta(k \log((n+1)/k))$ , and we propose a scheme (named ADMOT) that uses  $\mathcal{O}(k \log((n+1)/k))$  time slots<sup>1</sup>. Note that in each time slot, every transmitter transmits one symbol. Thus, one time slot is also one *symbol duration*.

Note that the general network scenario is also studied in which there are *multiple* transmitters, relay nodes and receivers. Again, the scaling law of estimating all network channels is achieved in a *reliable, computational efficient* and *distributed* manner.

#### A. Illustrating Example and Background Ideas

Consider a toy network consisting of three transmitting nodes  $\{S_1, S_2, S_3\}$  and one receiving node  $R$ . The three channels  $(S_1, R)$ ,  $(S_2, R)$  and  $(S_3, R)$  need to be estimated. Without loss of generality, let all the initial channel gains of the three channels be 1, and suppose one of the channel gains changes to  $x$  in the current time. The goal of monitoring is to identify the updated channel and the value of  $x$ . A simplistic monitoring scheme is to schedule transmissions on different channels in different time slots, as shown in Figure 1. In time slots 1, 2, and 3, sender  $S_i$ ,  $i = 1, 2, 3$ , sends probe signal 1 to node  $R$ , respectively, so that  $R$  can estimate the channel gain of  $(S_i, R)$ . Thus, altogether three time slots are needed.

However, using the algebraic approach to exploit the nature of wireless medium, **two** time slots are enough. As shown in Figure 2, in the first time slot  $S_1$  and  $S_2$  and  $S_3$  all send 1 to node  $R$ . These three signals “collide” in the air, but the collided signals turn out to be useful for our estimation. Let  $y[1]$  denote the signal received by  $R$  in the first time slot. We have  $y[1] = 3 + (x - 1)$ . In the second time slot  $S_1$ ,  $S_2$  and  $S_3$  send 1, 2 and 3, respectively.

<sup>1</sup> Note that  $\mathcal{O}(k \log((n+1)/k)) = \mathcal{O}(k \log(n/k))$ . In the paper we use  $\mathcal{O}(k \log((n+1)/k))$  to avoid the confusing case where  $k = n$ .

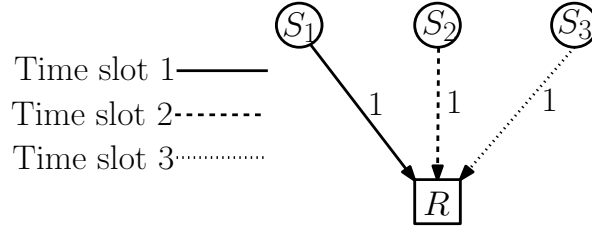


Fig. 1. The monitoring scheme based on scheduling. The “solid-line”, “dashed-line” and “dotted-line” are for the transmission of time slots 1, 2 and 3, respectively.

Thus, the received signal is  $y[2] = 6 + i(x - 1)$  if  $(S_i, R)$  is the updated channel. At the end of the second time slot,  $R$  computes  $[y(1), y(2)] - [3, 6] = (x - 1)[1, i]$ . Since  $[1, 1]$  and  $[1, 2]$  and  $[1, 3]$  are mutually independent,  $R$  can uniquely decode  $i$  and  $x$ .

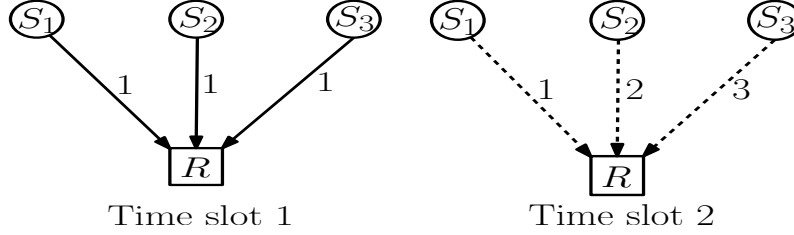


Fig. 2. A better monitoring scheme. The first and second sub-figure show the transmissions in time slots 1 and 2, respectively.

We summarize the main ingredients that give rise to the above saving as follows:

**(a) Embracing Interference for Group Probing.** In group probing, all the  $n$  channels are probed simultaneously in each time slot. This is essential to get rid of the  $\Theta(n)$  overhead in the traditional unit probing in which the channels are probed one by one. Note the  $\Theta(n)$  overhead in unit probing is fundamental even if the number of channels suffering appreciable variations,  $k$ , is much smaller than  $n$ . This is because we do not know which channels have changed.

**(b) Algebraic Distinguishability.** With respect to the illustrating example in Figure 2, for each  $i \in \{1, 2, 3\}$ , the training data of  $S_i$  (i.e.,  $\{1, i\}$ ) induces an “*algebraic fingerprint*” for channel  $(S_i, R)$ . Due to the linear independence of the fingerprints, the one corresponding to the varied channel is not erasable even under wireless interference. Note that it is not necessary to construct independent fingerprints by increasing the probing power. Later Section III shows that probing data with uniform magnitude but random signs suffice.

## B. Overview of Our Results

For the scenario where a receiver wants to monitor the channel gains from  $n$  transmitters, we first prove that the lower-bound of overhead is  $\Theta(k \log((n + 1)/k))$  time slots, where  $k \leq n$  is the number of channels suffering appreciable channel gain variations since the last round estimation. Note that when  $k = n$ , the overhead lower-bound

is  $\mathcal{O}(n)$ .

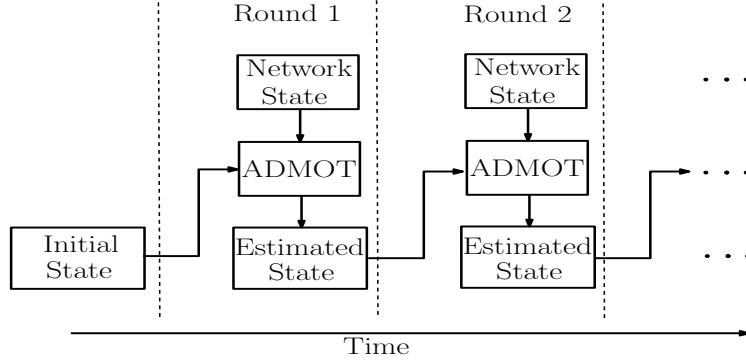


Fig. 3. Systematical implementation of ADMOT.

We then comprehensively develop the ideas in the above toy example and propose a scheme called Algebraic Differential SNR-Monitoring (ADMOT), which allows the receiver in the network to *reliably* monitor the channels with *minimum* overhead. For a systematical view, Figure 3 shows ADMOT in action for successive rounds<sup>2</sup>. In this figure, the *network state* is the set of channel gains of all channels, and the initial state is one in which every channel has zero channel gain. For the current *round* of monitoring, ADMOT estimates the network state from the estimation of the previous round. In the following, we summarize the desirable features of ADMOT.

- 1) ADMOT is *optimal*. For any  $k \leq n$ , ADMOT allows the receiver to *reliably* estimate all  $n$  channels within  $\mathcal{O}(k \log((n+1)/k))$  time slots, which matches the lower bound  $\Theta(k \log((n+1)/k))$ . To our best knowledge, all previous monitoring results require  $\Theta(n)$  regardless of  $k$ . Thus, ADMOT significantly reduces the overhead (compared with previous schemes) for small  $k$ , and preserves optimal performance even when  $k$  is close to  $n$ . These arguments are also verified by simulation in Section VI.
- 2) Under ADMOT, the computational complexity of the receiver is dominated by *convex optimization programming*, which can be computed in an *efficient* manner [14].
- 3) ADMOT is a *feedback-free* monitoring scheme, *i.e.*, the receiver does not need to send feedback during the monitoring and no centralized controller is assumed. Thus, the probing data can be incorporated as packet-header for practical packet transmissions [1], [5], [9].
- 4) ADMOT supports different *modulations* [2] at the physical layer. For example, BPSK could be used. That is, the sources can code the training data (of ADMOT) into binary symbols for BPSK modulation, such that both channel attenuation and phase term can be estimated.

We note that the *single receiver scenario* is the *fundamental* setting for studying the scaling law of monitoring wireless network. In Section V, the above results are applied to achieve the scaling law of general communication networks in which there are *multiple* transmitters, relay nodes and receivers.

<sup>2</sup>The interval between two monitoring rounds depends on the statistics of the channel coherence time and channel stability requirement.

### C. Related Work

Previous works fall into the following two categories.

**(a) Channel monitoring in wireless networks.** The works [8], [9], [10], [11], [12] designed probing data and estimation algorithm for estimating channel gains, and the works [15], [3] proposed schemes to estimate channel interference. In the first set of works (which are related to our work), interference has not been shown to be an advantage (compared with *nonoverlapping* probing signals by different transmitters), and the overhead achieved is  $\Theta(n)$ . Note that in the domain of wireless *network coding communication*, the work [4] was the first to show the advantage of interference, and later the work [5] proposed an amplify-and-forward relaying strategy for easy implementation.

**(b) Compressive sensing for channel estimation.** ADMOT proposed in this paper uses recent advances of compressive sensing developed for sparse signal recovering [16], [17]. Compressive sensing was used to recover the sparse features of channels, say channel's delay-Doppler sparsity [18], [19], channel's sparse multipath structure [20], [19], sparse-user detection [21], [22], [23] and channel's sparse response [24]. When applying above schemes to estimate all the  $n$  channels from the transmitters, the overhead is at least  $\Theta(n)$ . In contrast, ADMOT uses compressive sensing to handle all channels' *differential* information (embedded in the overlapped probing) simultaneously, and achieves optimal overhead  $O(k \log((n+1)/k))$ .

Note that some previous schemes mentioned above estimate the property of a wideband channel, in which the channel gain varies across the frequency within the channel bandwidth. In contrast, this paper investigates the scaling law (on the network size) of wireless network monitoring. For the sake of exposition, we focus on narrowband channels in which the channel gain is flat across the bandwidth of the channel. We believe that within the same scaling law complexity, ADMOT can easily be generalized to OFDM systems [2], in which information is carried across multiple narrowband channels.

### D. Organization of the paper

The rest of this paper is organized as follows. Section II formulates the problem. The scaling law theorem and the construction of ADMOT are presented in Section III. In Section IV, ADMOT is implemented by BPSK modulation. In Section V, monitoring in general communication networks is studied. Experimental results are shown in Section VI to support the theoretical analysis of ADMOT.

## II. PROBLEM SETTING AND PRELIMINARIES

### A. Notation Conventions and Preliminaries

Let  $\mathbb{Z}$  be the set of all integers and  $\mathbb{Z}^+$  be the set of all positive integers. Let  $\mathbb{R}$  be the set of all real numbers (*i.e.*, the real field). For any  $a$  and  $b$  in  $\mathbb{Z}^+$ , let  $\mathbb{R}^{a \times b}$  be the set of all matrices with dimensions  $a \times b$  and components

chosen from  $\mathbb{R}$ . For any matrix  $M \in \mathbb{R}^{a \times b}$  and  $i \in \{1, 2, \dots, a\}$  and  $j \in \{1, 2, \dots, b\}$ , let  $M(i, j)$  be the  $(i, j)$ 'th component of  $M$ .

Let  $\mathbb{R}^a$  be the set of all vectors with length  $a$  and components chosen from  $\mathbb{R}$ . For any vector  $V \in \mathbb{R}^a$  and  $i \in \{1, 2, \dots, a\}$ , let  $V(i)$  be the  $(i)$ 'th component of  $V$ . *Vectors in the paper are in the column form.*

For any vector  $V \in \mathbb{R}^n$ , let  $\|V\|_1 = \sum_{i=1}^n |V(i)|$  denote the  $\ell_1$ -norm and  $\|V\|_2 = \sqrt{\sum_{i=1}^n |V(i)|^2}$  denote the  $\ell_2$ -norm. For any vector  $V \in \mathbb{R}^n$  and non-negative integer  $k \leq n$ , we define the “distance” between  $V$  and  $k$ -sparsity by:

$$d_k(V) = \|V - V^k\|_1, \quad (1)$$

where  $V^k$  is  $V$  with all but the largest  $k$  components set to 0. Vector  $V$  is said to be  $k$ -sparse if and only if  $d_k(V) = 0$ , that is there are at most  $k$  nonzero components in  $V$ .

Let  $\mathbb{C}$  be the set of all complex numbers. The matrices and vectors over  $\mathbb{C}$  has the similar definitions. For any scalar, vector or matrix  $X$  over  $\mathbb{C}$ , let  $Re(X)$  be the real part of  $X$  and  $Im(X)$  be the imaginary part of  $X$ . For vector  $H \in \mathbb{C}^n$ , let  $\|V\|_2^2 = \|Re(H)\|_2^2 + \|Im(H)\|_2^2$ .

Let  $\mathcal{N}(\mu, \sigma^2)$  denote the normal distribution over *real field*, where  $\mu$  is the *mean* and  $\sigma^2$  is the variance. Note that throughout the paper, the logarithm function  $\log(\cdot)$  is computed over base 2, *i.e.*,  $\log(\cdot) = \log_2(\cdot)$ .

### B. Communication Model for Wireless Network

Consider a network where  $\mathcal{S} = \{S_1, S_2, \dots, S_n\}$  is the set of transmitting nodes. We first consider the scenario where there is only one receiving node  $R$ .

We assume all transmissions are slotted and synchronized. In each time slot, each and every transmitter transmits one symbol. Thus, one time slot is also one symbol duration and slotted synchronization is the same as symbol synchronization as well. Assume each  $S_i \in \mathcal{S}$  transmits symbol  $X_i[s] \in \mathbb{C}$  in time slot  $s$ . Thus, for time slot  $s$  the received signal at  $R$  is

$$Y[s] = \sum_{i=1}^n H(i)X_i[s] + Z[s], \quad (2)$$

where  $H(i) \in \mathbb{C}$  is the channel gain of  $(S_i, R)$  and  $Z[s] \in \mathbb{C}$  is the noise. Note that both  $Re(Z[s])$  and  $Im(Z[s])$  are *identically and independently distributed* (i.i.d.)  $\sim \mathcal{N}(0, 1)$  across all time slots. The **state** of  $R$  is defined to be a vector  $H \in \mathbb{C}^n$ , whose  $i$ 'th component is  $H(i)$ .

The transmit power and noise power are both *normalized* to be equal to 1 and the amplitude of the channel gain  $|H(i)|$  is the square-root of the signal-to-noise ratio (SNR) of  $(S_i, R)$ . For instance, for channel  $(S_i, R)$ , assume  $G(i)$  is the “true” channel gain,  $P$  is the transmit power and  $\sigma^2$  is the noise power. Therefore,  $\text{SNR}(S_i, R) = P|G(i)|^2/\sigma^2$ . After normalizing the transmit power and noise power, the amplitude of channel gain is  $|H(i)| = |G(i)|\sqrt{P}/\sigma =$

$$\sqrt{\text{SNR}(S_i, R)}.$$

As noted in Section I-B, such *single receiver scenario* is the basic setting for understanding the scaling law of wireless network monitoring. In Section V, we consider communication networks with *multiple* transmitters, relay nodes and receivers.

### C. Variation in Wireless Network

Wireless network conditions vary with time due to fading, transmit power instability, environmental changes and movement of mobile wireless devices. For the purpose of communication, network variation is *mathematically* equivalent to the variation of the state  $H$  (the definition of  $H$  can be found in Section II-B). Let  $\hat{H} \in \mathbb{C}^n$  be the *a priori* knowledge of previous state held by the receiving node  $R$ , and  $H$  be the current state. The monitoring objective of  $R$  is to estimate  $H$  using  $\hat{H}$  and the received probes. **Note** if  $R$  has no *a priori* knowledge of the previous network state,  $\hat{H}$  is set to be the zero vector in  $\mathbb{C}^n$ . For  $\epsilon > 0$  and non-negative integer  $k \leq n$ , the difference  $H - \hat{H}$  is said to be  $(k, \epsilon)$ -sparse if and only if

$$d_k(\text{Re}(H - \hat{H})) \leq \epsilon \quad \text{and} \quad d_k(\text{Im}(H - \hat{H})) \leq \epsilon,$$

where function  $d_k(\cdot)$  is defined in Equation (1).

For the simplicity, when  $\epsilon$  is small, “ $(k, \epsilon)$ -sparse” is also said to be “approx- $k$ -sparse”. Thus, for “approx- $k$ -sparse” variation  $H - \hat{H}$ , there are at most  $k$  channels suffering from *significant* variations for the channel gains, while the variations of other channels are negligible. In the following section ADMOT is proposed to catching up with the the  $k$  major varied channels with minimum overheads.

## III. ACHIEVING THE SCALING LAW BY ADMOT

In this section, a *novel* wireless network monitoring protocol ADMOT is proposed to achieve the scaling law shown in Theorem 1. To reduce the overhead of wireless monitoring, ADMOT fully develops the motivations shown in Section I-A. Furthermore, ADMOT exploits recent advances in the field of compressive sensing ([16], [17]), such that its *correctness* and *optimality* can be *theoretically* proved. A systematical view of ADMOT for consecutive wireless network monitoring can be found in Figure 3.

### A. Training Data of ADMOT

The training data of ADMOT is denoted by matrix  $\Phi$  with dimensions  $N \times n$ . Here,  $n$  is the number of transmitters in the network and  $N$  is the upper-bound of time slots used by ADMOT. Each component  $\Phi(s, i)$  is generated independently from  $\{-1, 1\}$  with equal probability, for all  $s$  and all  $i$ . The  $i$ -th column of matrix  $\Phi$  is assumed to

be known *a priori* to transmitter  $S_i$ , for all  $i \in \{1, 2, \dots, n\}$ . The knowledge of  $\Phi$  can be broadcast by  $R$  in the network setting stage<sup>3</sup>.

The training data of each  $S_i \in \mathcal{S}$  (i.e., the  $i$ 'th column of  $\Phi$ ) is in fact the ‘‘algebraic fingerprint’’ of channel  $(S_i, R)$ . As the toy example shown in Figure 2, these fingerprints are ‘‘highly independent’’ such that the varied channels would expose their fingerprints even under interference. In the next subsection, using a convex program, ADMOT can catch up with the exposed fingerprints in an efficient and reliable manner.

### B. Complete Construction of ADMOT

Before the detailed construction of ADMOT, a *convex-optimization* problem is proposed which serves as a submodule for ADMOT.

- **ConvexOPT**( $A, Y, \sigma$ ). The input of ConvexOPT( $A, Y, \sigma$ ) is  $A \in \mathbb{R}^{m \times n}$  and  $Y \in \mathbb{R}^m$  and  $\sigma > 0$ . The output of ConvexOPT( $A, Y, \sigma$ ) is the solution  $X^* \in \mathbb{R}^n$  to the following problem:

$$\min \|X\|_1 \quad \text{subject to} \quad \|AX - Y\|_2 \leq \sigma. \quad (3)$$

Note that ConvexOPT( $A, Y, \sigma$ ) is a *second-order cone programming* and can be solved efficiently [14].

Let  $m \leq N$  be the system parameter denoting the number of time slots used by the current round of ADMOT. We construct:

- **ADMOT**( $\hat{H}, \mathcal{S}, R, m$ ).
- variables Initialization: Vector  $H^* \in \mathbb{C}^n$  is the estimation of  $H$ , which is initialized to be zero vector. Vector  $Y \in \mathbb{C}^m$  is initialized to be zero vector. Let  $\Phi_m$  be the matrix consisting of the 1, 2, ...,  $m$ 'th rows of  $\Phi$ .
- Step A: For  $s = 1, 2, \dots, m$ , in the  $s$ 'th time slot:<sup>4</sup>
  - For any  $S_i \in \mathcal{S}$ ,  $S_i$  sends  $\Phi(s, i)$ .
  - Node  $R$  sets  $Y(s)$  (i.e., the  $s$ 'th component of  $Y$ ) to be the received sample in the time slot. Thus,  $Y(s) = \sum_{i=1}^n \Phi(s, i)H(i) + Z(s)$ , where  $Z(s)$  is the noise in the time slot (see Section II-B for details).
- Step B: Node  $R$  computes  $D \in \mathbb{C}^m$  as  $D = Y - \Phi_m \hat{H}$ . Thus,  $D = \Phi_m(H - \hat{H}) + Z$ .
- Step C: Node  $R$  runs ConvexOPT( $\Phi_m, \text{Re}(D), \sqrt{2m}$ ) and ConvexOPT( $\Phi_m, \text{Im}(D), \sqrt{2m}$ ). Let the solutions be denoted by  $\text{Re}(\Delta^*) \in \mathbb{R}^n$  and  $\text{Im}(\Delta^*) \in \mathbb{R}^n$ , respectively.
- Step D: Node  $R$  estimates  $H$  by  $H^* = \hat{H} + \Delta^*$ .

<sup>3</sup> To avoid the overhead of broadcasting  $\Phi$ , we can generate  $\Phi$  by practical pseudorandom generators (such as AES [25]). To be concrete, the  $i$ 'th column of  $\Phi$  could be the output of AES( $i$ ). Thus, each node in the network can compute  $\Phi$  using AES. Note that since ADMOT can be simulated within polynomial time, pseudo randomness suffices [25].

<sup>4</sup>Note that the probing scheme of ADMOT looks like CDMA [2]. For the clarification we note the difference between ADMOT and CDMA as: 1) CDMA is for information data detection, while ADMOT is for channel estimation. 2) CDMA requires near-orthogonal code sequence for each transmitter. In ADMOT, since  $m$  could be much less than  $n$ , the training data sent by each  $S_i \in \mathcal{S}$  can be far from orthogonal. However, the combination of ADMOT and CDMA is an interesting direction for future research.

- Step E: **End** ADMOT( $\hat{H}, \mathcal{S}, R$ ).

Thus, ADMOT can be performed in a feedback-free manner, *i.e.*, the receiver does not need to send feedback during the monitoring and no centralized controller is assumed. Under ADMOT, the computational complexity of the receiver node  $R$  is dominated by running the second-order cone program ConvexOPT, which can be solved in an efficient manner [14].

### C. Main theorem of the paper

The *main theorem* of the paper is:

*Theorem 1: ADMOT is optimal.*

- **Scaling law.** For any  $k \leq n$ , when  $H - \hat{H}$  is  $(k, \delta\sqrt{k})$ -sparse, any monitoring scheme achieving estimation error  $\|H^* - H\|_2 \leq \mathcal{O}(\delta)$  requires at least  $\Theta(k \log((n+1)/k))$  time slots.
- **Achievability.** Let  $k > 0$  be the maximum integer satisfying  $C_0 k \log((n+1)/k) \leq m$  for a constant  $C_0$ , and  $\delta > 0$  be the minimum real number such that  $H - \hat{H}$  is  $(k, \delta\sqrt{k})$ -sparse. The estimation error of ADMOT satisfies  $\|H^* - H\|_2 \leq \sqrt{2}C_1\delta + 2C_2$  with a probability  $1 - \mathcal{O}(e^{-0.15m})$ . Here,  $C_0$ ,  $C_1$  and  $C_2$  are constants defined in Appendix A.

The detailed proof for the theorem is in Appendix B. We have the following remarks regarding the theorem.

**Remark 1:** When  $H - \hat{H}$  is approx- $k$ -sparse (*i.e.*,  $\delta$  is small), the theorem shows that ADMOT reliably estimates  $H$  within  $\mathcal{O}(k \log((n+1)/k))$  time slots, which achieves the scaling law.

**Remark 2:** Recall that we normalize both probe power and noise power for the clarity. As shown in Section II-B, the “true” channel gain  $G(i)$  is in fact  $G(i) = H(i)\sigma/\sqrt{P}$ , where  $P$  is the transmit power and  $\sigma^2$  is the noise power. Thus, the “true” channel gains are estimated by  $G^* = \sigma H^*/\sqrt{P}$ . When  $G - \hat{G}$  is  $(k, \delta_G\sqrt{k})$ -sparse,  $H - \hat{H}$  is  $(k, \delta\sqrt{k})$ -sparse with  $\delta = \delta_G\sqrt{P}/\sigma$ . Thus, the estimation error of  $G$  is then  $\|G^* - G\|_2 \leq (\sqrt{2}C_1\delta + 2C_2)\sigma/\sqrt{P} = \sqrt{2}C_1\delta_G + 2C_2\sigma/\sqrt{P}$ . Thus, for large probing power  $P$ , the error term  $2C_2\sigma/\sqrt{P}$  caused by noise disappears and  $\|G^* - G\|_2$  approximates  $\sqrt{2}C_1\delta_G$ .<sup>5</sup>

### D. Adjusting system parameter $m$

The system parameter  $m$  corresponds to the trade-off between overheads and estimation errors. Ideally, we should choose  $m = C_0 k \log((n+1)/k)$ , where  $k$  is the number of channel gains which suffer significant variations since the last round estimation. Thus, the receiver  $R$  should estimate the *typical* number of varied channels between two monitoring rounds, and then adjust  $m$  for future rounds of ADMOT<sup>6</sup>.

<sup>5</sup>Note that for the case where each node in  $\mathcal{S}$  has different probing power, similar argument can be shown with somewhat unwieldy notations.

<sup>6</sup>Note that the receiver  $R$  can inform its choosing of  $m$  to other nodes by broadcasting before the next round of ADMOT. For instance, consider a cellular network where the receiver is the base station [1]. The information of  $m$  can be delivered in the stage of *downlink* transmission.

This object can be achieved by analyzing the estimation error in the past rounds. To be concrete, consider a past round of ADMOT with system parameter  $m$ . Let  $Y$  be the received data in this round. Receiver  $R$  divides  $Y \in \mathbb{C}^m$  into two parts (one for “estimation” and the other for “testing”): Vector  $Y_1 \in \mathbb{C}^{m-d}$  comprises of the first  $m-d$  components of  $Y$  and  $Y_2 \in \mathbb{C}^d$  comprises of the last  $d$  components of  $Y$ . Similarly, matrix  $\Phi_{(m,1)}$  comprises of the first  $m-d$  rows of  $\Phi_m$  and  $\Phi_{(m,2)}$  comprises of the last  $d$  rows of  $\Phi_m$ .

Receiver  $R$  runs Step B, C and D of ADMOT by using  $Y_1$  and  $\Phi_{(m,1)}$  (instead of  $Y$  and  $\Phi_m$ , respectively). Let  $H_t^*$  be the estimation of  $H$ . Let  $D_2 = Y_2 - \Phi_{(m,2)}H_t^*$  and  $\|H - H_t^*\|_2 = \varphi$ . Then we have:

*Theorem 2:* The event  $\|D_2\|_2^2 > d(\varphi\sqrt{3/2} + 2)^2$  happens with a probability at most  $O(e^{-0.15d})$ . If  $\varphi > 2\sqrt{2}$ , the event  $\|D_2\|_2^2 < d(\varphi/\sqrt{2} - 2)^2$  happens with a probability at most  $O(e^{-0.15d})$ .

The proof is delivered into Appendix C.

The theorem shows that the estimation error  $\|H - H_t^*\|$  preserves a close relationship with  $\|D_2\|_2$ . Thus, when  $\|D_2\|_2$  is large,  $R$  concludes that  $m-d$  time slots do not suffice to estimate  $H$ . Note that since  $d$  is relatively small (compared with  $m$ ),  $m$  time slots probing is also not reliable for estimating  $H$ . Thus,  $R$  should **increase**  $m$  for future rounds of ADMOT.

On the other hand, small  $\|D_2\|_2$  implies  $m-d$  time slots are sufficient for estimating  $H$ . For precisely estimating the minimum time slots which suffice,  $R$  can update  $d$  to  $2d$  and then re-computes  $\|D_2\|_2$ .

Then,  $R$  repeats this process until it finds the minimum integer  $p$  such that  $m - pd$  time slots are not sufficient. In the end, based on  $pd$ ,  $R$  can choose an appropriate **decreasing** of  $m$  for future rounds of ADMOT.

#### IV. BPSK IMPLEMENTATION OF ADMOT

In this section, we describe ADMOT implementation using BPSK modulation [2]. That is, we assume the symbols in  $\Phi$  are BPSK symbols.

Each node  $S_i \in \mathcal{S}$  transmits a BPSK symbol  $\Phi(s, i) \in \{-1, 1\}$  in time slot  $s$ . All transmitters transmit their symbols on the angular carrier frequency  $\omega$ . Let  $T$  be the duration of a time slot. Thus, in continuous time,  $S_i$  transmits the signal  $x_i(t) = \text{Re}(\sum_s \Phi(s, i)p(t - sT)e^{j\omega t})$ , where  $p(t) = 1$  for  $0 \leq t < T$ ; and  $p(t) = 0$  otherwise. Let the channel gain associated with  $S_i$  be  $H(i) = A_i e^{-j\theta_i}$ , where  $A_i \in [0, +\infty)$  is the amplitude of the channel gain and  $\theta_i$  is the phase delay due to signal propagation delay.

The signal reaching  $R$  from  $S_i$  is then  $\text{Re}(\sum_s \Phi(s, i)p(t - sT)e^{j\omega t} A_i e^{-j\theta_i})$ . Taking into the consideration the signals from all nodes in  $\mathcal{S}$  and the circuit noise, the combined signal at  $R$  is

$$y(t) = \sum_i \text{Re}\left(\sum_s \Phi(s, i)p(t - sT)e^{j\omega t} A_i e^{-j\theta_i}\right) + z(t),$$

where  $z(t)$  is the noise.

We assume  $T \gg 1/\omega$ . By matched filtering (i.e., multiplying  $y(t)$  by  $\cos(\omega t)$  and integrating over successive symbol periods  $T$ ; and multiplying  $y(t)$  by  $\sin(\omega t)$  and integrating over successive symbol periods), we can get

$$Y_{\cos}[s] = \sum_i A_i \cos(\theta_i) \Phi(s, i) + \frac{2}{T} \int_{(s-1)T}^{sT} z(t) \cos(\omega t) dt,$$

$$Y_{\sin}[s] = \sum_i A_i \sin(\theta_i) \Phi(s, i) + \frac{2}{T} \int_{(s-1)T}^{sT} z(t) \sin(\omega t) dt.$$

Note that the power of noise are normalized to 1 such that  $\frac{2}{T} \int_{(s-1)T}^{sT} z(t) \cos(\omega t) dt$  and  $\frac{2}{T} \int_{(s-1)T}^{sT} z(t) \sin(\omega t) dt$  are both i.i.d.  $\sim \mathcal{N}(0, 1)$ .

Thus, there are two set of channels and no inter-set inference happens. The state of the first set of channels is  $H_{\cos} \in \mathbb{R}^n$ , where the  $i$ 'th component is  $H_{\cos}(i) = A_i \cos(\theta_i)$ . And state of the second set of channels is  $H_{\sin} \in \mathbb{R}^n$ , where the  $i$ 'th component is  $H_{\sin}(i) = A_i \sin(\theta_i)$ .

Following ADMOT, the receiver can estimate  $H_{\sin}$  and  $H_{\cos}$  simultaneously (as monitoring  $Re(H)$  and  $Im(H)$ , see Section III-B). Once  $H_{\sin}$  and  $H_{\cos}$  are estimated,  $\{(A_i, \theta_i) : S_i \in \mathcal{S}\}$  can be computed efficiently. To be concrete, let  $H_{\sin}^*$  and  $H_{\cos}^*$  be the estimations of  $H_{\sin}$  and  $H_{\cos}$ , respectively. Thus,  $\theta_i$  can be estimated by  $\theta_i^* = \tan^{-1}(H_{\sin}^*(i)/H_{\cos}^*(i))$  and  $A_i$  can be estimated by  $A_i^* = \sqrt{(H_{\cos}^*(i))^2 + (H_{\sin}^*(i))^2}$ .

Note that ADMOT can also be implemented with other modulations. Due to the limit of space, we only present the BPSK implementations for ADMOT in this paper.

## V. ACHIEVING THE SCALING LAW FOR GENERAL COMMUNICATION NETWORK

In this section, we study the general communication network with multiple transmitters, receivers and intermediate relay nodes, e.g., the cooperative communication networks [26]. To be concrete, we can model the general communication network as  $(\mathcal{S}, \mathcal{R}, \mathcal{C})$ , where  $\mathcal{S} = \{S_1, S_2, \dots, S_n\}$  is set of transmitters,  $\mathcal{R} = \{R_1, R_2, \dots, R_{n'}\}$  is the set of receivers and  $\mathcal{C} = \{C_1, C_2, \dots, C_{n''}\}$  is the set of intermediate nodes for relaying. Let  $\mathcal{E}$  denote the set of all the channels:  $\{(S_i, R_j), (C_a, R_j), (S_i, C_a), (C_a, C_b) : S_i \in \mathcal{S}, R_j \in \mathcal{R}, C_a, C_b \in \mathcal{C}\}$ . Each channel in  $\mathcal{E}$  is either used for communication or considered as a interfering channel, and therefore needs to be monitored.

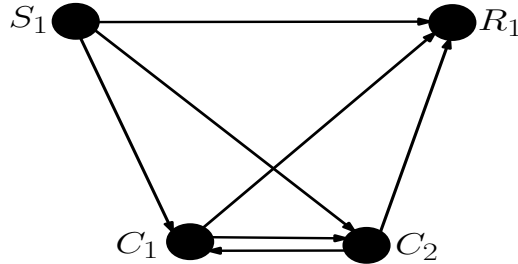


Fig. 4. The channels in an illustrating communication network  $(\mathcal{S}, \mathcal{R}, \mathcal{C})$ . In the figure,  $\mathcal{S} = \{S_1\}$ ,  $\mathcal{R} = \{R_1\}$  and  $\mathcal{C} = \{C_1, C_2\}$ . The directed lines denote the channels which are used for communications and therefore require monitoring.

For each node  $\beta \in \mathcal{R} \cup \mathcal{C}$ , let  $H_\beta \in \mathbb{C}^{n+n''}$  be the channel gains of the channels from  $\mathcal{S} \cup \mathcal{C}$  to  $\beta$ , and  $\hat{H}_\beta \in \mathbb{C}^{n+n''}$  be the *a priori* knowledge of these channel gains preserved by  $\beta$ . Assuming for each node in  $\mathcal{R} \cup \mathcal{C}$ , its state variation  $H_\beta - \hat{H}_\beta$  is approx- $k$ -sparse for  $k \leq n + n''$ . Using the scaling law in Theorem 1, at least  $\Theta(k \log((n + n'' + 1)/k))$  time slots are needed.

Under the full-duplex model, where any node in  $\mathcal{C}$  can transmit and receive in the same time slot,  $\text{ADMOT}(\hat{H}_\beta, \mathcal{S} \cup \mathcal{C}, \beta, m)$  can be performed simultaneously for each  $\beta \in \mathcal{R} \cup \mathcal{C}$ . Thus, we can achieve above overhead lower bound by choosing  $m = \Theta(k \log((n + n'' + 1)/k))$ .

For the half-duplex model, any node in  $\mathcal{C}$  cannot transmit and receive in a same time slot. Note that in the half-duplex model, for any node  $C_j \in \mathcal{C}$  the channel gain of  $(C_j, C_j)$  is always assumed to be 0. In next subsection we propose a *non-straightforward* generalization of ADMOT (named ADMOT-GENERAL) to achieve the scaling law  $\Theta(k \log((n + n'' + 1)/k))$ .

Note that for both models, our achievable schemes can be performed in a distributed manner, *i.e.*, no centralized controller is needed.

#### A. ADMOT-GENERAL

To handle the half-duplex feature, for each time slot ADMOT-GENERAL randomly selects (with  $1/2$  probability) nodes in  $\mathcal{C}$  to send probe data, while other nodes in  $\mathcal{C}$  receive signal in the time slot. To be concrete, let  $\Phi \in \mathbb{R}^{N \times (n+n'')}$  be the training data matrix. Each component of  $\Phi$  is i.i.d. chosen from  $\{0, -1, 1\}$  with a probability  $\{1/2, 1/4, 1/4\}$ . For each  $S_i \in \mathcal{S}$ , the  $i$ 'th column of  $\Phi$  is the training data of  $S_i$ . For each  $C_j \in \mathcal{C}$ , the  $n + j$ 'th column of  $\Phi$  is the training data of  $C_j$ . In ADMOT-GENERAL, for the  $s$ 'th time slot, if the training data of  $C_j$  is zero,  $C_j$  would receive signal in the time slot; Otherwise  $C_j$  would send the corresponding probe data.

Choosing  $m = 3C'_0 k \log(n/k) = \Theta(k \log(n/k))$  where  $C'_0$  is a constant defined in Appendix A, we have:

- **ADMOT-GENERAL**( $m$ )<sup>7</sup>.
- **variables Initialization:** For each  $\beta \in \mathcal{R} \cup \mathcal{C}$ , vector  $H_\beta^* \in \mathbb{C}^{n+n''}$  is the estimation of  $H$ , which is initialized to be zero vector. For each  $\beta \in \mathcal{R} \cup \mathcal{C}$ , let  $\mathcal{I}_\beta = \{i : \Phi(i, n + j) = 0, i \in \{1, 2, \dots, m\}\}$  if  $\beta = C_j \in \mathcal{C}$ , and  $\mathcal{I}_\beta = \{1, 2, \dots, m\}$  if  $\beta \in \mathcal{R}$ . For each  $\beta \in \mathcal{R} \cup \mathcal{C}$ , let  $m_\beta = |\mathcal{I}_\beta|$ , and  $\Phi_\beta \in \mathbb{R}^{m_\beta \times (n+n'')}$  consist of the rows of  $\Phi$  which are indexed by  $\mathcal{I}_\beta$ , vector  $Y_\beta \in \mathbb{C}^{m_\beta}$  be initialized to be zero vector.
- **Step A:** For  $s = 1, 2, \dots, m$ , in the  $s$ 'th time slot:
  - For any  $S_i \in \mathcal{S}$ ,  $S_i$  sends  $\Phi(s, i)$ .
  - For any  $C_j \in \mathcal{C}$ , if  $\Phi(s, n + j) = 0$ ,  $C_j$  receives signal in this time slot; Otherwise  $C_j$  sends  $\Phi_\beta(s_\beta, n + j)$ , where  $s_\beta \in \mathcal{I}_\beta$  is the index of the row in  $\Phi_\beta$  which corresponds to the  $s$ 'th row of  $\Phi$ .

<sup>7</sup>For the simplicity, we omit other parameters  $(\{\hat{H}_\beta : \beta \in \mathcal{S} \cup \mathcal{C}\}, \mathcal{S}, \mathcal{C}, \mathcal{R})$ .

- Any node in  $\mathcal{R}$  receives signal in this time slot.
- For each node  $\beta \in \mathcal{R} \cup \mathcal{C}$ , if  $\beta$  received signal in this time slot,  $\beta$  sets  $Y_\beta(s_\beta)$  to be the received sample.

Thus,  $Y_\beta(s_\beta) = \sum_{i=1}^{n+n''} \Phi_\beta(s_\beta, i) H_\beta(i) + Z_\beta(s)$ , where  $Z_\beta(s)$  is the noise in the time slot.

- Step B: For each  $\beta \in \mathcal{R} \cup \mathcal{C}$ ,  $\beta$  computes  $D_\beta \in \mathbb{C}^{m_\beta}$  as  $D_\beta = Y_\beta - \Phi_\beta \hat{H}_\beta$ . Thus,  $D_\beta = \Phi_\beta(H_\beta - \hat{H}_\beta) + Z_\beta$ , where  $Z_\beta \in \mathbb{C}^{m_\beta}$  is the noise vector for  $\beta$ .
- Step C: For each  $\beta \in \mathcal{R} \cup \mathcal{C}$ ,  $\beta$  runs  $\text{ConvexOPT}(\Phi_\beta, \text{Re}(D_\beta), \sqrt{2m_\beta})$  and  $\text{ConvexOPT}(\Phi_\beta, \text{Im}(D_\beta), \sqrt{2m_\beta})$ .<sup>8</sup> Let the solutions be denoted by  $\text{Re}(\Delta_\beta^*) \in \mathbb{R}^{n+n''}$  and  $\text{Im}(\Delta_\beta^*) \in \mathbb{R}^{n+n''}$ , respectively.
- Step D: For each  $\beta \in \mathcal{R} \cup \mathcal{C}$ ,  $\beta$  estimates  $\beta$  by  $H_\beta^* = \hat{H}_\beta + \Delta_\beta^*$ .
- Step E: **End** ADMOT-GENERAL.

For each  $\beta \in \mathcal{R} \cup \mathcal{C}$ , using Chernoff Bound [27], we have  $m_\beta \geq m/3 \geq C'_0 k \log(n/k)$  with a probability at least  $1 - 2^{-m}$ . Using the achievability result in Theorem 1, we conclude  $\beta$  can recover  $H_\beta$  with bounded square root errors<sup>9</sup>.

## VI. PERFORMANCE EVALUATION

We evaluate ADMOT by implementing it in a systematical manner, as shown in Figure 3. Let the  $n = |\mathcal{S}| = 500$ , and the average channel SNR= 20db.

Recall that a channel is said to preserve *stability*  $x\%$  if the probability is no more than  $(1 - x\%)$  that the channel suffers significant variation during the interval between two monitoring rounds.

In the simulation, let  $H[r]$  be the state of the  $r$ 'th round. Thus, for the channel stability  $x\%$ ,  $H[r] = H[r-1] + \Delta[r]$ , where  $\Delta[r]$  is the variation. Each component of  $\Delta[r] \in \mathbb{C}^n$ , say  $\Delta[r](i)$ , is *independently* generated as: With a probability  $x\%$ , both  $\text{Re}(\Delta[r](i))$  and  $\text{Im}(\Delta[r](i))$  are uniformly chosen from  $[-10, 10]$ ; With a probability  $1 - x\%$ , both  $\text{Re}(\Delta[r](i))$  and  $\text{Im}(\Delta[r](i))$  are uniformly chosen from  $[-250, 250]$ .

We proceed  $\text{ADMOT}(H^*[r-1], \mathcal{S}, R, m_r)$  for the  $r$ 'th round estimating. Here,  $H^*[r-1]$  is the estimation of  $H[r-1]$  in the  $(r-1)$ 'th round, and the system parameter  $m_r$  is chosen depending on the receiving data in the previous rounds (see Section III-D for details).

Figure 5 shows the average time slots (per round) used by ADMOT for  $x \in (0, 100)$ . In the figure, the solid line is for ADMOT, and the dashed line is for previous monitoring schemes (see related works in Section I-C). From the figure, we can see that ADMOT significantly reduces the overheads for the scenarios where  $x$  is large, *i.e.*, high channel stability is required. In the region where  $x$  is small, ADMOT also preserves reliable performance.

<sup>8</sup>Note that if  $\beta = C_j \in \mathcal{C}$ , since the channel gain of  $(C_j, C_j)$  is always assumed to be zero, the  $n + j$ 'th components of  $\text{Re}(D_\beta)$  and  $\text{Im}(D_\beta)$  are both fixed to be zero for running ConvexOPT.

<sup>9</sup>Note that  $\Phi_\beta$  satisfies RIP of order  $k$  with a probability at least  $1 - 2^{-n}$  (see Appendix A), which is the sufficient requirement for applying Theorem 1 (see the proof in Appendix B).

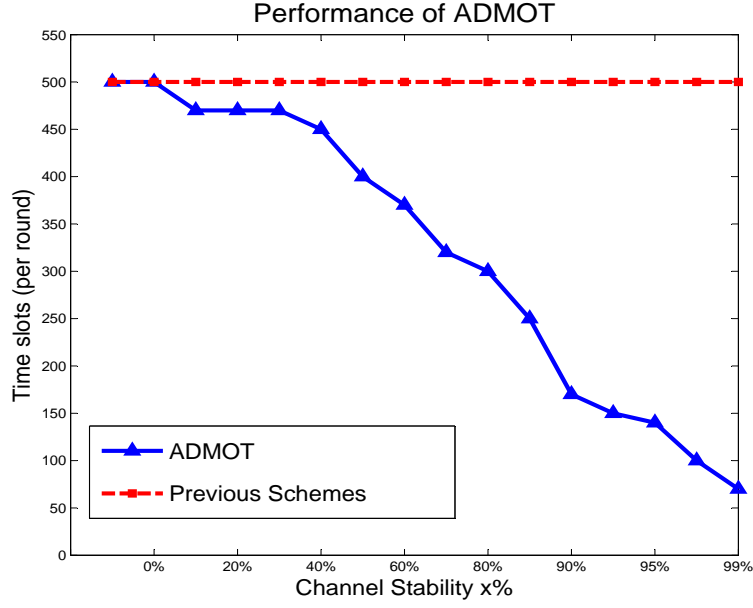


Fig. 5. Comparison between ADMOT and previous monitoring schemes.

We also provide the detailed simulations for the cases where channel preserves stabilities 80%, 90%, and 98%, respectively.

Figure 6 shows the time slots used by ADMOT for round  $r \in \{1, 2, \dots, 50\}$ . For the channel stability 80%, 90%, and 98%, the average time slots used per round are 320, 252, and 140, respectively. Note that since we assume zero knowledge for the initial network state, the first round of each case costs almost 500 time slots.

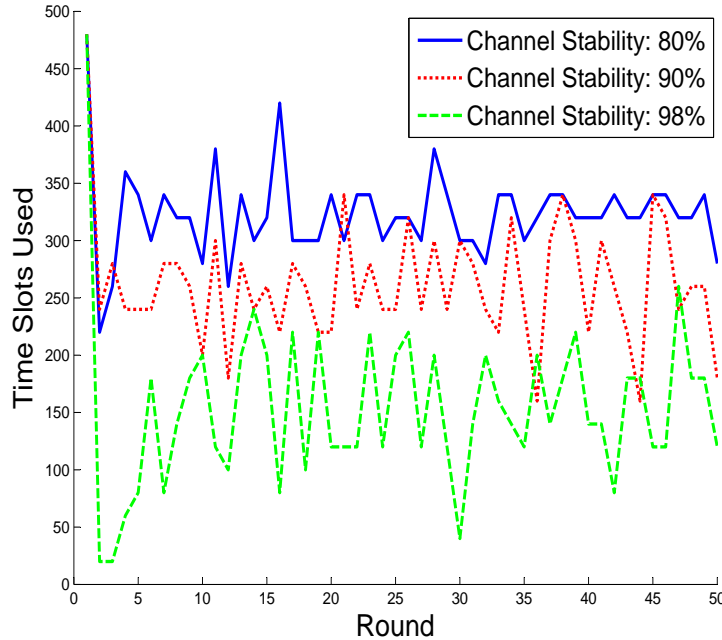


Fig. 6. Estimating time for each stage of running ADMOT.

Figure 7 shows the relative estimation errors ( $\|H^*[r] - H[r]\|_2 / \|H[r]\|_2$ ) of ADMOT for round  $r \in \{1, 2, \dots, 50\}$ . Note that we bound estimation error regardless the channel stability  $x\%$ . Thus, lower channel stability only corresponds to more overheads (as shown in Figure 6).

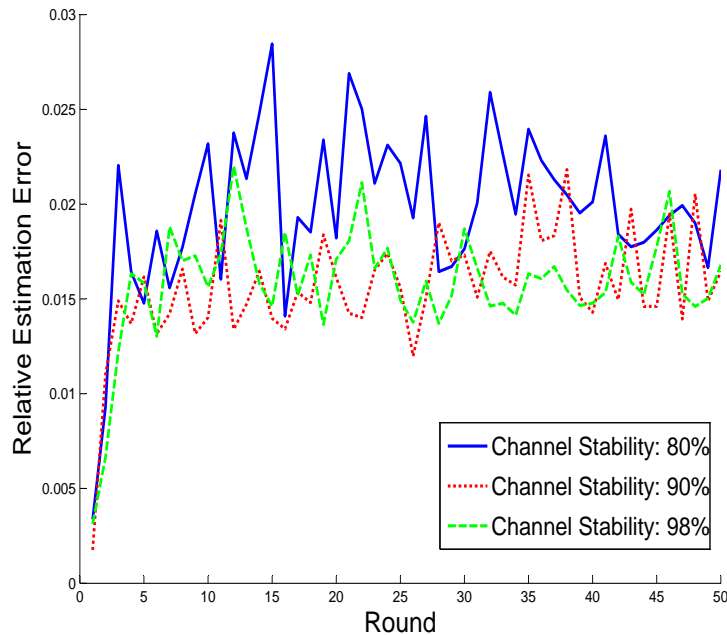


Fig. 7. Relative estimating error for each stage of running ADMOT.

For a detailed looking, we also show the estimations at round 50 for the case of 80% channel stability. Figure 8 draws (the absolute value of the real part) channel gains and the corresponding estimations for the 200, 201, ..., 300'th channels.

## VII. CONCLUSION

In the paper, we first investigate the scenario where a receiver needs to track the channel gains of the channels with respect to  $n$  transmitters. We assume that in current round of channel gain estimation, no more than  $k \leq n$  channels suffer significant variations since the last round. We prove that “ $\Theta(k \log((n+1)/k))$  time slots” is the *minimum* number of time slots needed to catch up with  $k$  varying channels. At the same time, we propose ADMOT to achieve the lower bound in a computationally efficient manner. Furthermore, ADMOT supports different *modulations* at the physical layer. Using above results, we also achieve the scaling law of general communication networks in which there are multiple transmitters, relay nodes and receivers. In the end of the paper, we also present simulation results to support our theoretical analysis.

## REFERENCES

- [1] D. Shen, Z. Pan, K. K. Wong, and V. O. Li, “Effective throughput: A unified benchmark for pilot-aided ofdm/sdma wireless communication systems,” in *Proc. of INFOCOM*, 2003.

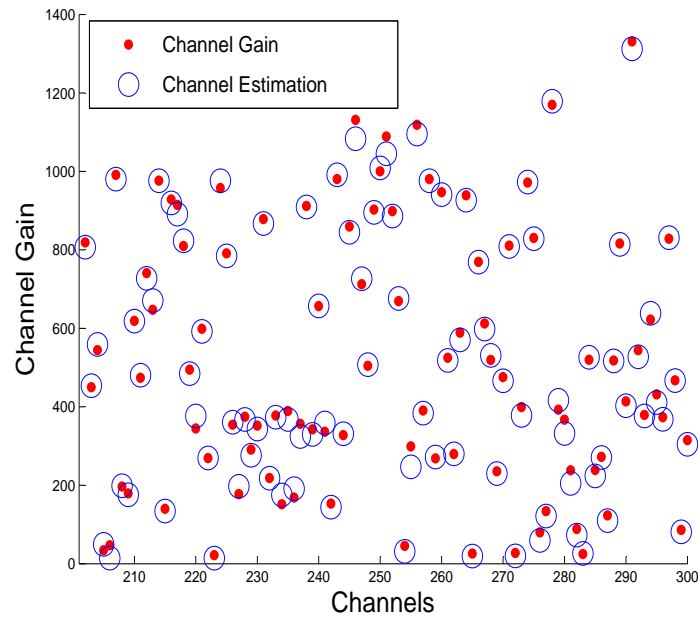


Fig. 8. Detailed drawing of channel gains and the corresponding estimations.

- [2] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [3] K. Ramachandran, E. Belding-Royer, K. Almeroth, and M. Buddhikot, "Interference-aware channel assignment in multi-radio wireless mesh networks," in *Proc. of IEEE INFOCOM*, 2006.
- [4] S. Zhang, S.-C. Liew, and P. Lam, "Physical-layer network coding," in *Proc. of MOBICOM*, 2006.
- [5] S. Katti, S. Gollakota, and D. Katabi, "Embracing wireless interference: Analog network coding," in *Proc. of SIGCOM*, 2007.
- [6] P. Sharma, S. J. Lee, J. Brassil, and K. G. Shin, "Distributed channel monitoring for wireless bandwidth aggregation," in *Proc. of IFIP-TC6 Networking Conference*, 2004.
- [7] J. Lee, S. J. Lee, P. Sharma, and S. Choi, "Understanding the effectiveness of a co-located wireless channel monitoring surrogate system," in *Proc. of IEEE ICC*, 2010.
- [8] E. Ertin, U. Mitra, and S. Siwamogsatham, "Maximum-likelihood based multipath channel estimation for code-division multiple-access systems," *IEEE Trans. Communications*, 1998.
- [9] A. Soysal and S. Ulukus, "Mimo multiple access channels with noisy channel estimation and partial csi feedback," in *Proc. of IEEE Globecom*, 2008.
- [10] S. Buzzi and H. V. Poor, "Channel estimation and multiuser detection in long-code ds/cdma systems," *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, vol. 19, no. 8, pp. 1476–1487, 2001.
- [11] M. K. Oh, X. Ma, G. B. Giannakis, and D. Park, "Cooperative synchronization and channel estimation in wireless sensor networks," *JOURNAL OF COMMUNICATIONS AND NETWORKS*, 2005.
- [12] X. Ma, M. K. Oh, G. B. Giannakis, and D. Park, "Hopping pilots for estimation of frequency-offset and multiantenna channels in mimo-ofdm," *IEEE Trans. on Communications*, vol. 53, no. 1, pp. 162–172, 2005.
- [13] G. Judd, X. Wang, and P. Steenkiste, "Efficient channel-aware rate adaptation in dynamic environments," in *Proc. of the 6th international conference on Mobile systems, applications, and services*, 2008.
- [14] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [15] J. Lee, S. J. Lee, W. Kim, D. Jo, T. Kwon, and Y. Choi, "Rss-based carrier sensing and interference estimation in 802.11 wireless

- networks,” in *Proc. of IEEE SECON*, 2007.
- [16] E. Candés, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
  - [17] D. Donoho, “Compressed sensing,” *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
  - [18] G. Tauböck and F. Hlawatsch, “A compressed sensing technique for ofdm channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots,” in *Proc. of ICASSP*, 2008.
  - [19] C. Berger, S. Zhou, J. Preisig, and P. Willett, “Sparse channel estimation for multicarrier underwater acoustic communication from subspace methods to compressed sensing,” *IEEE Trans. Signal Processing*, vol. 58, no. 3, pp. 1708–1721, March 2010.
  - [20] W. U. Bajwa, A. Sayeed, and R. Nowak, “Sparse multipath channels: Modeling and estimation,” in *Proc. of IEEE Digital Signal Proc. Workshop*, 2009.
  - [21] A. K. Fletcher, S. Rangan, and V. K. Goyal, “On-off random access channels: A compressed sensing framework,” *Arxiv submission: 0903.1022*, 2009.
  - [22] S. T. Qaseem1 and T. Y. Al-Naffouri, “Compressive sensing for reducing feedback in mimo broadcast channels,” in *Proc. of IEEE ICC*, 2010.
  - [23] S. R. Bhaskaran, L. Davis, A. Grant, S. Hanly, and P. Tune, “Downlink scheduling using compressed sensing,” in *Proc. of IEEE Information Theory Workshop (ITW)*, 2009.
  - [24] S. Cotter and B. Rao, “Sparse channel estimation via matching pursuit with application to equalization,” *IEEE Trans. on Communications*, vol. 50, no. 3, 2002.
  - [25] Y. Lindell and J. Katz, *Introduction to Modern Cryptography*. Chapman and Hall/CRC press, 2007.
  - [26] S. Sharma, Y. Shi, J. Liu, Y. T. Hou, and S. Kompella, “Is network coding always good for cooperative communications?” in *Proc. of INFOCOM*, 2010.
  - [27] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.
  - [28] E. Candés, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
  - [29] S. Mendelson, A. Pajor, and N. T. Jaegermann, “Uniform uncertainty principle for bernoulli and subgaussian ensembles,” *Constructive Approximation*, vol. 28, no. 3, pp. 277–289, 2008.
  - [30] K. D. Ba, P. Indyk, E. Price, and D. P. Woodruff, “Lower bounds for sparse recovery,” in *Proc. of SODA*, 2010.
  - [31] V. Vu, “Chernoff bound,” University of California, San Diego. Available at: <http://cseweb.ucsd.edu/~klevchen/techniques/chernoff.pdf>, Tech. Rep.

## APPENDIX A

### PRELIMINARIES OF COMPRESSIVE SENSING

Compressive sensing is a *mathematical* technique developed for compressible data recovering with significantly fewer samples than the length of data [16], [17]. All compressive sensing results used in the paper is introduced in this section.

Let  $M$  be a matrix in  $\mathbb{R}^{m \times n}$  with  $m \ll n$ . Assume each column of  $M$  is normalized to have  $\ell_2$ -norm 1. For positive integer  $k$ ,  $M$  is said to satisfy *restricted isometry property*(RIP) of order  $k$  if  $(1 - \delta_k) \|X\|_2^2 \leq \|MX\|_2^2 \leq (1 + \delta_k) \|X\|_2^2$  for all  $k$ -sparse vector  $X \in \mathbb{R}^n$  [28].

Let  $A \in \mathbb{R}^{m \times n}$  be a matrix. Then we have [29]:

- Each element of  $A$  is i.i.d. generated from  $\{-1, 1\}$  with equal probability  $\{1/2, 1/2\}$ . Then, with overwhelming probability (i.e.,  $1 - \mathcal{O}(2^{-n})$ ), matrix  $A/\sqrt{m}$  satisfies RIP of order- $k$  provided that  $m \geq C_0 k \log((n+1)/k)$ , where  $C_0$  is a constant depending on each instance.
- Each element of  $A$  is i.i.d. generated from  $\{0, -1, 1\}$  with probability  $\{1/2, 1/4, 1/4\}$ . Then, with overwhelming probability (i.e.,  $1 - \mathcal{O}(2^{-n})$ ), matrix  $\sqrt{2/m}A$  satisfies RIP of order- $k$  provided that  $m \geq C'_0 k \log((n+1)/k)$ , where  $C'_0$  is a constant depending on each instance.

Let  $X \in \mathbb{R}^n$  be the data vector and  $Y = AX + Z$  be the noisy measurement, where  $Z \in \mathbb{R}^m$  is the noise with  $\|Z\|_2 \leq \sigma$ . Let  $X^* \in \mathbb{R}^n$  be the solution to  $\text{ConvexOPT}(A, Y, \sigma)$ , where  $\text{ConvexOPT}(\cdot)$  is defined in Section III-B.

Assuming  $A/\sqrt{m}$  satisfies RIP, the following theorem is proved in [28].

*Theorem 3:* The solution  $X^*$  obeys

$$\|X - X^*\|_2 \leq C_1 d_k(X)/\sqrt{k} + C_2 \sigma/\sqrt{m}, \quad (4)$$

where  $C_1$  and  $C_2$  are constants and  $d_k(\cdot)$  is defined in (1).

## APPENDIX B

### PROOF FOR THEOREM 1

We show an intermediate lemma before the proof of Theorem 1. Let  $Z \in \mathbb{R}^m$  such that each of its component  $Z(i)$  is i.i.d.  $\sim \mathcal{N}(0, 1)$ . Then we have

*Lemma 4:* The  $\ell_2$ -norm  $\|Z\|_2 \leq \sqrt{2m}$  with a probability at least  $1 - e^{-0.15m}$ .

**Proof:** For any  $i \neq j$ ,  $Z(i)$  and  $Z(j)$  are independent and normally distributed. Then the probability density function of  $X = Z(i)^2 + Z(j)^2$  is  $f_X(x) = e^{-x/2}/2$  for  $x \geq 0$  [27]. Thus,  $E(e^{X/4}) = \int_0^{+\infty} e^{(-x/4)}/2 dx = 2$ .

Without loss of generality we assume ADMOT chooses  $m$  as an *even* integer. Then we have:

$$\begin{aligned} \Pr(\|Z\|_2^2 > 2m) &= \Pr\left(\sum_{i=1}^d Z(i)^2/4 > m/2\right) \\ &= \Pr\left(e^{\sum_{i=1}^m Z(i)^2/4} > e^{m/2}\right) \\ &\leq E\left(e^{\sum_{i=1}^m Z(i)^2/4}\right) / e^{m/2} && \text{Markov Inequality} \\ &= \prod_{j=1}^{m/2} E\left(e^{Z(2j-1)^2/4 + Z(2j)^2/4}\right) / e^{m/2} && \text{Independence} \\ &\leq 2^{m/2} / e^{m/2} \leq e^{-0.15m}. \end{aligned}$$

□

Then we have:

**Proof of the Scaling laws in Theorem 1:** Without loss of generality, we first consider a sub-problem (*i.e.*, an easier problem): Assuming  $Re(H - \hat{H})$  is  $(k, \sqrt{k}C)$ -variation and  $Im(H - \hat{H})$  is a all-zero vector, what is the minimum time slots required to find  $Re(H^*)$  such that  $\|Re(H^*) - Re(H)\|_2^2 = \mathcal{O}(1)$ ?

Assume  $T$  time slots are used for estimating  $Re(H)$ . For any  $s = 1, 2, \dots, T$ , and  $i = 1, 2, \dots, n$ , in the  $t$ 'th time slot let  $S_i$  send  $A(s, i) \in \mathbb{C}$ . Here  $A$  is a  $T \times n$  complex matrix whose  $(s, i)$ 'th component is  $A(s, i)$ .

Let  $Y(s) \in \mathbb{C}$  be the received data of  $R$  in the  $s$ 'th time slot, and therefore  $Y \in \mathbb{C}^T$  be a length- $T$  vector whose  $s$ 'th component is  $Y(s)$ . Thus  $Y = AH$ . Note that we assume that there is no noise here, which only reduces the complexity of estimating  $H$ .

Since  $\hat{H}$  and  $A$  are known by  $R$  as *a priori*, the original problem is equivalent to estimating  $\Delta = H - \hat{H}$  by  $D = Y - A\hat{H} = A(H - \hat{H})$ . Due to  $Im(H) = Im(\hat{H})$ , we have  $Y = A(Re(H) - Re(\hat{H}))$ . Thus the problem is equivalent to estimating  $Re(\Delta)$  by using  $Re(D)$  and  $Im(D)$ , which compose of  $2T$  linear samples (over  $\mathbb{R}$ ) of  $Re(\Delta)$ .

A recent result [30] in the field of compressive sensing proves that provided  $d_k(Re(\Delta)) \leq C\sqrt{k}$  for some constant  $C$ , it requires at least  $\Theta(k \log((n+1)/k))$  linear samples (over  $\mathbb{R}$ ) for reliably finding  $\Delta^* \in \mathbb{R}^n$  such that  $\|Re(\Delta) - \Delta^*\|_2^2 \leq \mathcal{O}(1)$ . Thus we have  $T \geq \Theta(k \log((n+1)/k))$ .

Thus, we prove the complexity of the easier problem. For the original problem, which considers random noise and the variations of imaginary parts of the channel gains, the complexity can only be higher.  $\square$

**Proof of the Achievability in Theorem 1:** Recall that the constant  $C_0$ ,  $C_1$ , and  $C_2$  are defined in Appendix A. Note that since  $m$  satisfies  $m \geq C_0 k \log((n+1)/k)$ , with overwhelming probability (*i.e.*,  $1 - \mathcal{O}(2^{-n})$ ), the matrix  $\Phi_m$  satisfies RIP of order- $k$  (Appendix A). We henceforth assume it is true.

We first analyze  $Re(D) = \Phi_m Re(H - \hat{H}) + Z$ , where  $Z \in \mathbb{R}^m$  is the noise term. Using Lemma 4, we have  $Pr(\|Z\|_2 \leq \sqrt{2m}) < e^{-0.15m}$ . We henceforth assume it is true.

Thus, using Theorem 3, vector  $Re(\Delta^*)$  satisfies  $\|Re(\Delta^*) + Re(\hat{H} - H)\|_2 \leq C_1 d_k(Re(\hat{H} - H))/\sqrt{k} + \sqrt{2}C_2$ .

Since the state variation  $H - \hat{H}$  is  $(k, \delta\sqrt{k})$ -sparse, after setting  $H^* = \hat{H} + \Delta^*$  we have  $\|Re(H^* - H)\|_2 \leq C_1\delta + \sqrt{2}C_2$ .

Similarly we have  $\|Im(H^* - H)\|_2 \leq C_1\delta + \sqrt{2}C_2$ . In the end, we have  $\|H^* - H\| \leq \sqrt{2}C_1\delta + 2C_2$ . It completes the proof.  $\square$

## APPENDIX C

### PROOF OF THEOREM 2

We first show an intermediate lemma. Let  $V \in \mathbb{R}^n$  be a fixed vector, and  $R \in \mathbb{R}^n$  be a vector of random variables. For each component  $R(i)$  we have  $Pr(R(i) = 1) = Pr(R(i) = -1) = 0.5$ , and  $R(i)$  is i.i.d. for all  $1 \leq i \leq n$ . Let  $\langle V, R \rangle = \sum_{i=1}^n V(i)R(i)$  denote the inner product between  $V$  and  $R$ .

**Lemma 5:** For  $\langle V, R \rangle^2$ , the expectation  $E(\langle V, R \rangle^2)$  is  $\|V\|_2^2$  and the variance  $\text{Var}(\langle V, R \rangle^2)$  is no more than  $\|V\|_2^4$ .

**Proof:** We have

$$\begin{aligned}\langle V, R \rangle^2 &= \sum_i R(i)^2 V(i)^2 + \sum_{i \neq j} R(i) R(j) V(i) V(j) \\ &= V(i)^2 + \sum_{i \neq j} R(i) R(j) V(i) V(j) \\ &= \|V\|_2^2 + \sum_{i \neq j} R(i) R(j) V(i) V(j).\end{aligned}$$

Since  $E(\sum_{i \neq j} R(i) R(j) V(i) V(j)) = 0$ , we have  $E(\langle V, R \rangle^2) = \|V\|_2^2$ .

We have:

$$\begin{aligned}E(\langle V, R \rangle^4) &= \|V\|_2^4 + E\left(\left(\sum_{i \neq j} R(i) R(j) V(i) V(j)\right)^2\right) \\ &= \|V\|_2^4 + E\left(\left(\sum_{i \neq j} R(j)^2 R(i)^2 V(j)^2 V(i)^2\right)\right) \\ &= \|V\|_2^4 + \left(\sum_{i \neq j} V(j)^2 V(i)^2\right) \\ &\leq \|V\|_2^4 + \left(\sum_i V(i)^2\right) \left(\sum_j V(j)^2\right) \\ &= 2\|V\|_2^4.\end{aligned}$$

Thus  $\text{Var}(\langle V, R \rangle^2) = E(\langle V, R \rangle^4) - E(\langle V, R \rangle^2)^2$  is no more than  $\|V\|_2^4$ .  $\square$

Then we have:

**Proof of Theorem 2:** Let  $\Delta = (H - H_1^*)$ ,  $\varphi_R = \|Re(\Delta)\|_2$ ,  $\varphi_I = \|Im(\Delta)\|_2$  and thus  $\varphi = \|\Delta\|_2 = \sqrt{\varphi_R^2 + \varphi_I^2}$ .

Without loss of generality, we first analyze  $\|Re(D_2)\|_2^2$ .

From the definition,  $Re(D_2) = U_R + Z_R$ , where each component of  $Z_R \in \mathbb{R}^d$  is i.i.d.  $\sim \mathcal{N}(0, 1)$  and  $U_R \in \mathbb{R}^d$  is  $\Phi_{(m,2)} \Delta_R$ . Using Lemma 5, for  $i = 1, 2, \dots, d$ , we have  $E(U_R(i)^2) = \varphi_R^2$  and  $\text{Var}(U_R(i)^2) \leq \varphi_R^4$ . Note that each component of  $\Phi$  is i.i.d.; the noise term is i.i.d.  $\sim \mathcal{N}(0, 1)$ ;  $U_R(i)$  and  $U_R(j)$  are independent for all  $i \neq j$ . Thus, we can apply Chernoff Bound (on discrete bounded random variables) [31] and get  $Pr\left(|\sum_{i=1}^d (U_R(i)^2 - \varphi_R^2)| / (n\varphi_R) > d\varphi_R^2 / (2n\varphi_R)\right) \leq 2e^{-d^2/16}$ . It is equivalent to

$$Pr(d\varphi_R^2/2 \leq \|U_R\|_2^2 \leq 3d\varphi_R^2/2) \geq 1 - 2e^{-d^2/16}$$

Using Lemma 4, we have  $Pr(\|Z_R\|_2^2 > 2d) \leq e^{-0.15d}$ .

Similarly, assuming  $Im(D_2) = U_I + Z_I$ , we have

$$\begin{aligned} Pr\left(d\varphi_I^2/2 \leq \|U_I\|_2^2 \leq 3d\varphi_I^2/2\right) &\geq 1 - 2e^{-d^2/16}, \\ Pr(\|Z_I\|_2^2 > 2d) &\leq e^{-0.15d}. \end{aligned}$$

Using Union Bound [27] for  $U_R$  and  $U_I$ , we have

$$Pr(d\varphi^2/2 \leq \|U_R + jU_I\|_2^2 \leq 3d\varphi^2/2) \geq 1 - 4e^{-d^2/16}.$$

Similarly, by the Union Bound for  $Z_R$  and  $Z_I$  we can derive

$$Pr(\|Z_R + jZ_I\|_2^2 > 4d) \leq 2e^{-0.15d}.$$

Note that  $D_2 = (U_R + jU_I) + (Z_R + jZ_I)$ . Using *triangle inequality*, the event  $\|D_2\|_2^2 > d(\varphi\sqrt{3/2} + 2)^2$  happens with a probability at most  $4e^{-d^2/16} + 2e^{-0.15d} = O(e^{-0.15d})$ .

Assuming  $\varphi^2 > 8$ , the event  $\|D_2\|_2^2 < d(\varphi/\sqrt{2} - 2)^2$  happens with a probability at most  $4e^{-d^2/16} + 2e^{-0.15d} = O(e^{-0.15d})$ . □